

Moving Towards Semantic Web: Relational Schema to Ontology

Asif Sohail Abid and Muhammad Younus Javed

Abstract—Over the past decade there is an increase in requirement of web application, and to the most of the web application relational schema is attached as a backend. The idea of semantic web was presented by Tim Berners-Lee, and it has captured the focus of many researchers in past few years. The promises of future web have brought the researchers to the point where to obtain the benefits of future web; we need to make conversion of current web i.e. web 2.0 to future Web i.e. Semantic Web. There will be relational schema to ontology conversion required while making such transitions. There is a need to have some automated or semi automated approach for this issue as it will be hard to convert relational schema to ontology manually. The paper discusses some of the available techniques for the conversion, and focuses on the foreign key identification in the relational schema then compares some already presented techniques and concludes the results on the experimentation carried out on different relational schemas.

Index Terms—Semantic web, R2O, analysis of the schemes for R2O.

I. INTRODUCTION

Semantic web is often named as future web. The idea of future web is presented by Tim Berners-Lee et al [1]. This web promises many features that are dramatically gaining the focus of many researchers in the past decades [2]. Semantic web aims to make the web understandable and meaningful.

The development in web technologies in past and requirement of the data intensive web pages that is mainly based on relational database are finding ways to migrate towards the semantic Web which is ontology based [3], [4]. The reverse engineering of relational schema to ontology is a tedious work if done manually, as there is different association and cardinalities to look after among the tables while conversion. The mapping and transformation are two different task as discussed by Astrova1 et al [5], in mapping it is assume that both ontologies and relational databases exists and produces set of correspondence between these two. Where as, on the other hand the transformation assumes only relational database exists and builds ontology from it. In many web applications the associations are maintain in the application layer of the website no at the backend database, so the analyzing of HTML form is also done for correct identification of association in relational schema.

The significance of this research paper is to check for the

available techniques for the conversion of relational schema to ontology, and find the association in the relational schema. We have also used those relational schemas where there is absence of metadata, and discussed different scenarios accordingly.

In organization of the paper is in such a way that, section-II presents some discussion on the related work, in section-III test schemas, effects of selecting thresholds, experimental evaluation and results are presented, and in section-IV gives the conclusion of the paper.

II. RELATED WORK

Ontology is one of the pillars of semantic web. It is logical, shareable and common understanding of the Domain. Irina Astrova et al [5] proposed a novel approach for the reverse engineering of database into ontologies by considering HTML forms analysis. A. Bucella, M. R. Penabad, F. J. Rodriguez, A. Farina, A. Cechich [6] proposes a method for generating OWL based initial source ontology to integrate data sources. It categorizes tables in two categories Tables without foreign keys and tables with foreign key.

M.li, X. Du, S. Wang [7] provides an ontology learning framework and group of learning rules for classes, properties, hierarchy, cardinalities, and instances to learn ontology from relational database. The technique assumes that all the tables are in third normal form.

Most of the efforts were made in which source code is used for the extraction of schema, and limited work exists on the reverse engineering of legacy data files [8], [9], [10]. The legacy data can be used to retrieve information like primary key, foreign key, and relations between tables. The following equations for foreign key in a table and relation between two tables are explaining the key rule for extracting these information.

$$type(PK_{TableA}) = Type(FK_{TableB}) \quad (1)$$

$$\forall V \in Values(FK_{TableB}) \Rightarrow V \in Values(PK_{TableA}) \quad (2)$$

The foreign key of Table B can be a primary key in Table A. it can also be concluded from the above equation that for all instances of Foreign Key of Table B can also be the instances of Primary key in Table A. We have selected some algorithms for identification of foreign key in different relational schemas. The following algorithm is from approach presented by Asif et al., [11]. Different weights have been used in the approach, which act as the tuning parameters for the identification of the associations.

Manuscript received September 10, 2012; revised November 21, 2012.

Asif Sohail Abid is with the Faculty of Computing, Riphah International University, Islamabad (e-mail: Asif.sohail@riphah.edu.pk).

Muhammad Younus Javed is with the College of Electrical and Mechanical Engineering, National University of Sciences and Technology, Rawalpindi, Pakistan (e-mail: myjaved@ceme.nust.edu.pk).

Inputs: The primary key (Pk) of a table and other fields of remaining tables (R')

```

Foreignkey_identification(Pk,[1,2,3,...,n])
i ← n; (n is numbers of remaining fields )
do
  R'_NAME ← int R'NAME[i];
  R'_DT ← int R'DT[i];
  if Pk_NAME == R'_NAME then
    weight ← weight + w1;
  if R'_NAME contains Pk_NAME then
    weight ← weight + w2;
  if R'_DT contains Pk_DT then
    weight ← weight + w3;
  i ← i - 1;
  while i != 0
return weight

```

Fig. 1. Algorithm-1.

III. EXPERIMENTAL EVALUATION AND RESULTS

The interested area in the evaluation is correctness in identification of association. For evaluation criteria system performance, preservation of information and correctness are considered. (a) System performance: It is the efficiency of the scheme in term of time. (b) Preservation of information capacity: This is to finds out whether the association in relational schema were completely identified or not. It is important to find information loss (if any) as a result of transformation. (c) Correct identification: This criterion evaluates the correctness of the transformation. The Table I shows system specification

TABLE I: SYSTEM SPECIFICATIONS.

SYSTEM REQUIREMENT	
RAM	1GB
Processor	Intel 1.37 Dual Core
Operating System	Windows XP

A. Test Schemas

There are three test schemas that were used for association identification. The first schema was selected from the paper asif et al. Table II shows some details of that schema.

The second relational schema was selected from paper “Extracting ontologies from relational databases”, whereas the third schema was a complex relational schema of a campus management system comprising of 57 tables and almost 27 possible relationships among tables.

TABLE II: DETAILS OF RELATIONAL SCHEMA.

ENTITY NAME	ATTRIBUTE
Office Code	Office code, City, phone, address Line1, Address Line2, State, Country, Postal Code, Territory
Employee	Employee Number, Last name, first name, Extension, email, office code, repost to, Job title
Orders	Order number, order Date, Required Date, Shipped date, Status, Comments, Customer Number, Employee Number
Customer Number	Customer Number, Customer Last Name, Customer First Name, Phone, addressLine1, AddressLine2,City
Product Line	Product Line, Text Description, Html Description, Image, Product Code
Product	Product Code, Product sale, Product Vendor, Product Description, Quantity in stock, buy price, order number
Order details	Order number, quantity ordered, price each, ordered Line Number
Payments	Customer Number, check number, parmentdate, amount

B. Selection of Weights for Algorithm-1

The algorithm-I is tested on the third relational schema. There are different tuning parameters for the algorithm as w_1 , w_2 and w_3 , as selection of weights for the identification of the association is sensitive task, which require domain knowledge. The different variations of weights are tested in order to show the effects on net result. Table III shows the effect of w_1 , w_2 and w_3 on the results.

TABLE III: SELECTION OF WEIGHTS.

IDENTIFIED	W1	W2	W3	FP	FN
28400	0	0.5	0.5	YES	0
24678	0.2	0.4	0.4	YES	0
16034	0.3	0.4	0.3	YES	0
5261	0.4	0.3	0.3	YES	0
5255	0.5	0.3	0.2	YES	0
29	0.6	0.3	0.2	0	0
29	0.7	0.2	0.1	0	0
29	0.8	0.1	0.1	0	0
27	0.9	0.1	0	0	YES
27	1	0	0	0	YES
30	0.5	0	0.5	YES	0
29	0.4	0.1	0.4	0	0
30	0.4	0.2	0.3	YES	0
30	0.3	0.3	0.3	YES	0
30	0.3	0.4	0.2	YES	0
5254	0.3	0.5	0.2	YES	0
5254	0.2	0.6	0.1	YES	0
5254	0.1	0.7	0.1	YES	0
5254	0.1	0.8	0	YES	0
5254	0	0.9	0.1	YES	0
5254	0	1	0	YES	0
28400	0	0.5	0.5	YES	0
24678	0.2	0.4	0.4	YES	0
16034	0.3	0.4	0.3	YES	0
5261	0.4	0.3	0.3	YES	0
5255	0.5	0.3	0.2	YES	0
29	0.6	0.3	0.2	0	0
29	0.7	0.2	0.1	0	0
29	0.8	0.1	0.1	0	0
27	0.9	0.1	0	0	0
27	1	0	0	0	YES

where identified are the number of association identified in the transaction with weights w_1 , w_2 and w_3 . FN represents the false negative and FP represents the false positive values in the operation. From the above table we have concluded that if the weight w_1 , w_2 and w_3 are kept in this scenario as

$$0.6 \leq w_1 \leq 0.8 \quad (3)$$

$$w_2 = 0.4 \quad (4)$$

$$0.6 \leq w_3 \leq 0.9 \quad (5)$$

In the Fig. 2 the Table II is shown graphically. The values of weights are not fixed it varies from one test schema to another.

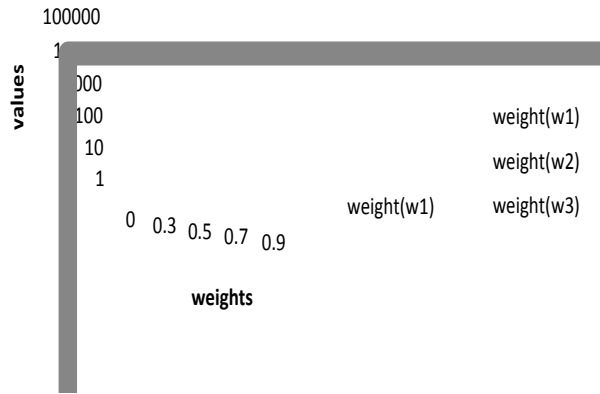


Fig. 2.

IV. PRECISION AND RECALL

The technique by Asif et al is compared with Abbasifard et al and Guohua Shen et al schemes. Both techniques that are considered for the comparison are selected because of their addressing to same problem to which proposed solution refers. The following graphs in Fig. 3 and Fig. 4 shows the precision and recall of some schemes on relational schema.

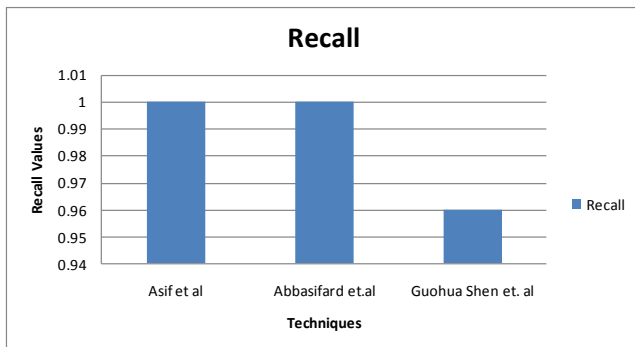


Fig. 3. Recall graph.

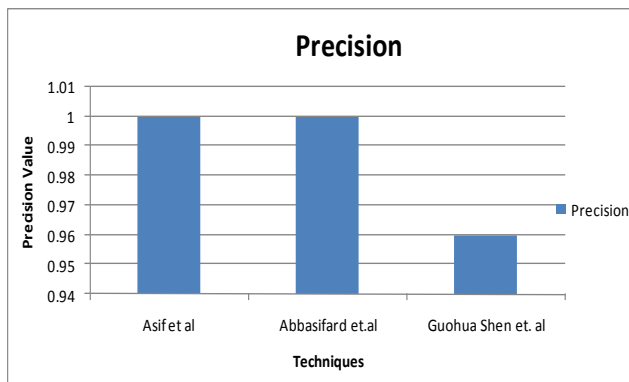


Fig. 4. Precision graph.

A. Time Complexity Graph

The Asif et al., technique is compared with abbasifard et al and Guohua Shen et al schemes. The Fig. 5 shows the time required graph of the schemes used in identification of Associations.

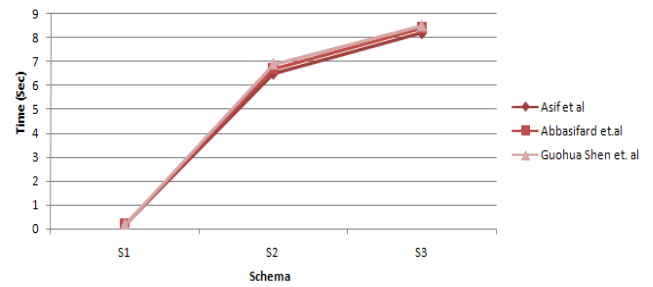


Fig. 5.

V. CONCLUSION

There are different techniques available for the conversion of relational schema to ontology. The Precision and recall of asif et al., and Abbasifard et al are good , the time of these three schemes are different as some techniques uses more rules for transformation and some uses less.

REFERENCES

- [1] B. L. Tim, J. Hendler, and O. Lassila, "The semantic web," *Scientific American Magazine*, Retrieved March 26, 2008.
- [2] M. R. Abbasifard, M. Rahgozar, A. Bayati, and P. Pornemati, "Using automated database reverse engineering for database integration," in *Proceeding of World Academy of Science, Engineering and Technology*, 2006.
- [3] L. Stojanovic, N. Stojanovic, and R. Volz, "Migrating data-intensive web sites into the semantic web," in *Proceedings of the 17th ACM Symposium on Applied Computing (SAC)*, pp. 1100-1107, 2002.
- [4] L. Lubyte and S. Tessaris, "Extracting ontologies from relational databases," in *Proceedings of the 20th Int. Workshop on Description Logics*, 2007.
- [5] I. Astroval and B. Stantic, "Reverse engineering of relational databases to ontologies: An approach based on an analysis of HTML forms," in *Proceedings of the 23rd IASTED International Conference on Databases and Applications (DBA)*, Innsbruck, Austria. pp. 246-251, 2005.
- [6] A. Bucella, M. R. Penabad, F. J. Rodriguez, A. Farina, and A. Cechich "From relational databases to OWL ontologies," *Digital Libraries: Advanced Methods and Technologies. Digital Collection*, pages-Puschchino, Rusia, 2004.
- [7] M. Li, X. Du, and S. Wang, "Learning ontology from relational databases," *Machine learning and Cybernetics*, vol. 6, pp. 3410-3415, 2005.
- [8] M. Casanova and D. S. Amaral, "Mapping uninterpreted schemes into entity-relationship diagrams, two applications to conceptual schema design," *IBM J. Res. & Dev.*, vol. 28, no. 1, 1984.
- [9] K. Davis and A. Arora, "A methodology for translating a conventional file system into an entity-relationship model," in *Proceedings of ERA, IEEE/North-Holland*, 1985.
- [10] P. Fraternali, "Tools and approaches for developing data-intensive web applications," *A Survey, in ACM Computing Surveys*, vol. 31, no. 3, pp. 227-263, 1999.
- [11] A. S. Abid and M. Y. Javed, "An effective approach for association identification in real scenarios," *Second International Conference on Computer and Electrical Engineering*, 2009.
- [12] M. Laclav'ik, "RDB2Onto: Relational database data to ontology individuals mapping," in *Proceeding of Ninth international Conference of Informatics*, 2007.
- [13] M. Casanova and D. S. Amarel, "Designing entity relationship schemas for conventional information systems," in *Proc. of Entity-Relationship Approach*, pp. 265-278, 1983.